Guided Selection of Accurate Belief Propagation Fixed Points

Christian Knoll
Signal Processing and Speech Comm. Lab.
Graz University of Technology
christian.knoll.c@ieee.org

Florian Kulmer
Signal Processing and Speech Comm. Lab.
Graz University of Technology
florian_kulmer@gmx.at

Franz Pernkopf
Signal Processing and Speech Comm. Lab.
Graz University of Technology
pernkopf@tugraz.at

Abstract

Belief propagation (BP) and the Bethe approximation are two closely related concepts that both suffer from the existence of multiple fixed points (or stationary points). We propose a modification of BP, termed self-guided belief propagation (SBP), that incorporates the pairwise potentials only gradually; this essentially selects one specific fixed point and increases the accuracy without increasing the computational burden. We apply SBP to various models with Ising potentials and show that: (i) SBP is superior in terms of accuracy whenever BP converges, and (ii) SBP obtains a unique, stable, and accurate solution whenever BP does not converge.

1 Introduction

Probabilistic graphical models provide a general framework to represent high-dimensional distributions. Two fundamental problems that arise in this context are: computing the marginal distributions and evaluating the partition function. Both problems are NP-hard to solve for models that contain loops and thus require efficient approximation methods [3]. Belief propagation (BP) exploits the structure of the model in an efficient way but may fail to converge, or has its approximation accuracy suffer from the existence of multiple fixed points.

Over the years a fruitful connection between computer science and concepts from statistical physics emerged [22, 13]. In particular, the relationship between fixed points of BP and stationary points of the Bethe free energy $\mathcal{F}_\mathcal{B}$ led to a deeper understanding of the underlying properties [5, 14]. This spurred the development of a whole class of algorithms that operate on $\mathcal{F}_\mathcal{B}$ directly, including methods that aim to obtain or approximate [16, 19, 21] the global minimum of the (possibly non-convex) Bethe free energy. Besides being non-trivial to minimize, the global minimum of $\mathcal{F}_\mathcal{B}$ may even be sub-optimal in terms of marginal accuracy [9]. This motivates the pursuit for methods that overcome these issues, enhance the convergence properties, and reliably provide accurate approximations [11, 4].

In this work, we present self-guided belief propagation (SBP) that aims to fill this gap by building upon the observation that strong interactions reduce accuracy and deteriorate the convergence properties [8]. SBP first considers only local potentials (where BP is exact and has a unique solution) and subsequently modifies the model by increasing the pairwise potentials to the desired values. This deterministic sequence of models thus iteratively refines the Bethe approximation towards an accurate solution.

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

We evaluate SBP for grid-graphs and random graphs with Ising potentials. Compared to BP, we observe superior performance in terms of accuracy in a fraction of runtime. SBP further excels for models where BP fails to converge and provides accurate marginals nonetheless.

2 Background

We consider a finite set of N discrete random variables $\mathbf{X} = \{X_1, \ldots, X_N\}$ that take values $x_i \in \mathcal{X} = \{-1, +1\}$. The graphical model $\mathcal{U} = (\mathcal{G}, \Psi)$ is then defined by a set of potentials Ψ and by an undirected graph $\mathcal{G} = (\mathbf{X}, \mathbf{E})$, where the random variables are in a one-to-one correspondence with the nodes and where $\mathbf{E}$ is the set of edges. For each node $X_i \in \mathbf{X}$ we denote the set of neighbors by $\partial(i) = \{X_j \in \mathbf{X} : (i,j) \in \mathbf{E}\}$. In this work, we focus on pairwise models that play an important role in various scientific fields. We further assume that all possible configurations have non-zero probability, which admits an exponential factorization. Now, let the pairwise potentials $\Phi(x_i, x_j) = \exp(J_{ij} x_i x_j)$ be defined by the couplings $J_{ij} \in \mathbb{R}$ that are assigned to each edge, and the local potentials $\Phi(x_i) = \exp(\theta_i x_i)$ be defined by the fields $\theta_i \in \mathbb{R}$ that act on each variable. Then the joint distribution factorizes according to

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\mathcal{Z}} \prod_{(i,j)\in\mathbf{E}} \Phi(x_i, x_j) \prod_{X_i \in \mathbf{X}} \Phi(x_i) = \frac{1}{\mathcal{Z}} \exp\Big(\sum_{(i,j)\in\mathbf{E}} J_{ij} x_i x_j + \sum_{i=1}^{N} \theta_i x_i \Big). \tag{1}$$

This particular model is known as Boltzmann machine in the machine learning community [21, 6] or as Ising model in the physics literature [2]. We distinguish two different types of interactions between random variables and call a model $\mathcal{U}$ *attractive* (or ferromagnetic) if all J_{ij} are positive, and refer to it as *general* model (or spin glass) if it contains both positive and negative J_{ij}.

BP provides an iterative way to efficiently obtain the marginal distributions $P(x_i)$ and the partition function $\mathcal{Z}$ for tree-structured graphs and to approximate these quantities for graphs that contain loops [10]. BP recursively exchanges messages between random variables according to

$$\mu_{ij}^{n+1}(x_j) \propto \sum_{x_i \in \mathcal{X}} \Phi(x_i, x_j) \Phi(x_i) \prod_{X_k \in \{\partial(i) \setminus X_j\}} \mu_{ki}^{n}(x_i). \tag{2}$$

The set of messages $\boldsymbol{\mu}^n$ contains the messages along all edges at iteration n; if all messages are converged we refer to the fixed point messages by $\boldsymbol{\mu}^\circ$. Then the approximate marginals are computed by $P(x_i) = \frac{1}{Z_i} \Phi(x_i) \prod_{X_k \in \partial(i)} \mu_{ki}^\circ(x_i)$. The marginals further define the Bethe free energy $\mathcal{F}_\mathcal{B}$ and the Bethe partition function $\mathcal{Z}_B = \exp(-\mathcal{F}_\mathcal{B})$ (cf. [18]) that approximates the partition function $\mathcal{Z}$.

3 Selecting an Accurate Fixed Point

BP often provides accurate results if it converges, but may also suffer from the existence of multiple fixed points of varying accuracy, which arise for models with loops that have sufficiently strong couplings. In this case the initial message values primarily determine to which fixed point BP will converge and thus play a crucial role in the overall performance. Unfortunately, the optimal initial values are model-dependent and are not known in general. Moreover, even if a set of fixed points is provided there is no way to tell which one has the most accurate marginals – even selecting the fixed point that minimizes $\mathcal{F}_\mathcal{B}$ may be non-optimal [9]. We introduce an iterative approach that revokes the role of the initialization and avoids above drawbacks.

3.1 Self-Guided Belief Propagation (SBP)

Inspired by the recent observation that strong local potentials increase accuracy and lead to better convergence properties [8], we aim to reduce the influence of the pairwise potentials that negatively influence BP. SBP starts from a simple model with independent random variables and slowly incorporates the potentials' strength, i.e., it solves the simple problem first and – by repetitive application of BP, keeps track of the fixed point as the interaction strength is increased by a scaling term.

In other words, SBP relaxes the problem of minimizing $\mathcal{F}_\mathcal{B}$ by making all variables independent (and the Bethe approximation exact). Then, the problem is deformed into the original one by increasing ζ from zero to one. Thereby, a stationary point ${\mathcal{F}_\mathcal{B}}^\circ$ emerges as a well-behaved path and SBP keeps

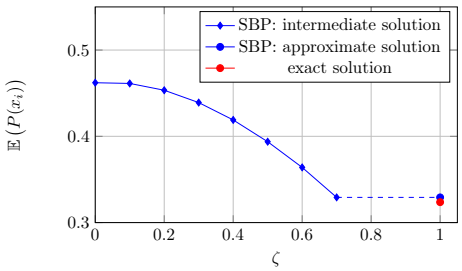


Figure 1: SBP proceeds along a smooth solution path and obtains accurate marginals despite instability of the terminal fixed point. Note that the fixed point becomes unstable for $\zeta > 0.7$; SBP stops and provides the last stable solution (that is already close to the exact one) as an approximation.

track of it with BP constantly correcting the stationary point.

More formally SBP considers an increasing length-M sequence $\{\zeta_m\}$ where $m = 1, \dots, M$ such that $\zeta_m < \zeta_{m+1}$ and $\zeta_m \in [0, 1]$ with $\zeta_1 = 0$ and $\zeta_M = 1$. This further indexes a sequence of probabilistic graphical models $\{\mathcal{U}_m\}$ that converges to the model of interest $\mathcal{U}_M = \mathcal{U}$. Every model has a set of potentials $\Psi_m = \{\Phi_m(x_i, x_j), \Phi_m(x_i)\}$ associated, where $\Phi_m(x_i) = \Phi(x_i)$ and the pairwise potentials at index m are exponentially scaled by $\Phi_m(x_i, x_j) = \exp(J_{ij}\zeta_m x_i x_j) = \Phi(x_i, x_j)^{\zeta_m}$. We further denote the fixed points of BP for $\mathcal{U}_m$ by $\boldsymbol{\mu}^\circ_m$. SBP provides a favorable initialization for every model $\mathcal{U}_m$ by the preceding fixed point $\boldsymbol{\mu}^\circ_{m-1}$.

This may lead to problems if the fixed point becomes unstable for some value $m < M$, in which case we cannot rely on BP to keep track of the fixed point anymore. Instead, SBP provides the last stable fixed point in that case, i.e., $\boldsymbol{\mu}^\circ_{m-1}$ as the final estimate. The path-tracking community breaks down such problems into three key steps: prediction, correction, and step-size adaption (cf. [1]). While the implementation details of these steps may influence the performance, we have found in our experiments that the overall efficiency of SBP remains largely unaffected.

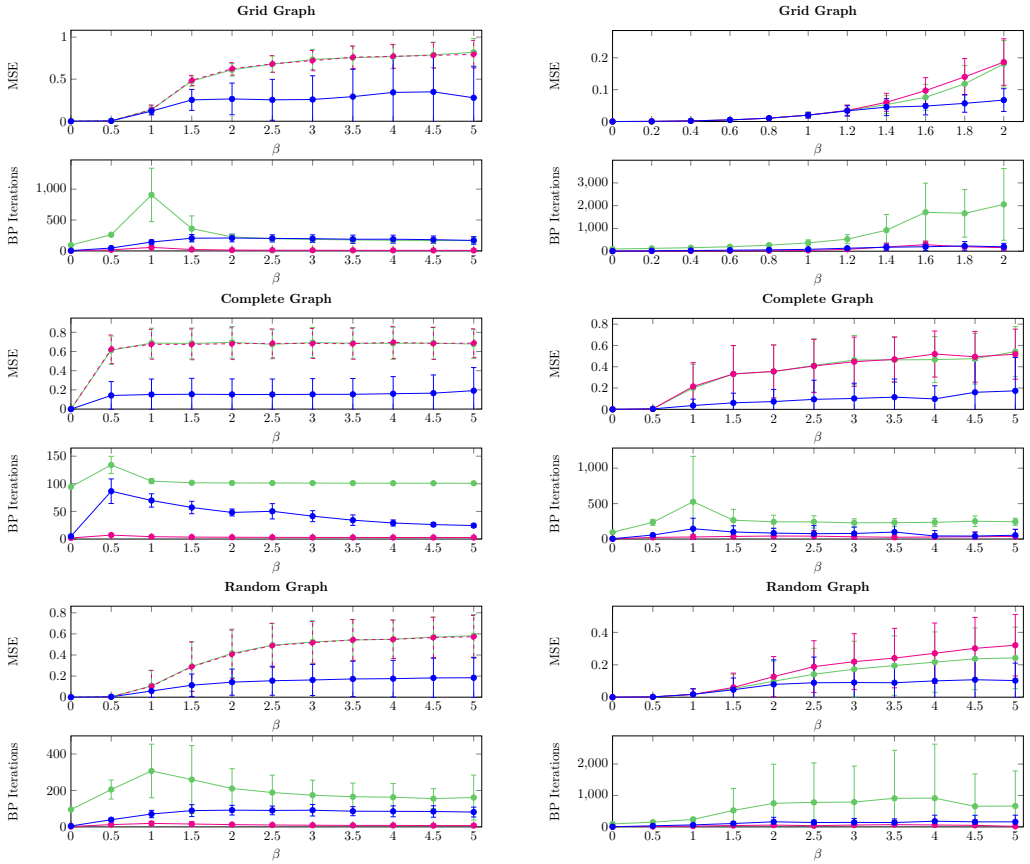
We illustrate how SBP approximates the marginals for a problem where BP fails to converge in Fig. 1. Initially SBP obtains the marginals for $\zeta = 0$ and then estimates the marginals of the desired problem by successively increasing ζ and running BP to keep track of the emerging solution path.

4 Experiments

We apply SBP to attractive and general models on 10×10 grid graphs, and on complete and random graphs with $N = 10$ variables. These graphs are considered in order to render the computation of the exact marginals feasible and to make the results comparable to previous work [20, 17, 12]. We evaluate SBP and compare it to BP and BP_D (BP with damping) [15] by the mean squared error (MSE) between the approximate marginals $\tilde{P}(x_i)$ and the exact marginals $P(x_i)$. We further compare the runtime by counting the overall number of BP iterations.

For BP and SBP we set the maximum number of iterations to $N_{BP} = 10^3$ and use random message scheduling. For BP_D we choose a large damping factor $\epsilon = 0.9$ in order to prioritize convergence over runtime and therefore increase the maximum number of iterations to $N_{BP} = 10^4$. The initial messages are randomly initialized 100 times for each model, before applying BP (and BP_D). We consider BP (and BP_D) as converged for a model if at least a single message initialization (out of 100) exists for which BP converges. The reported error (MSE) and the number of iterations are averaged over all convergent runs of BP and BP_D while all runs that did not converge are discarded. SBP, on the other hand, always provides an approximation, even if the terminal fixed point is unstable.

Note that the solution space of models with identical parameters $J_{ij} = J$ for all edges and $\theta_i = \theta$ for all nodes is well-structured for and specifically suits the application of SBP. In particular, the global minimum of the Bethe free energy provides the most accurate marginals (cf. [9, Lm.1]) and is continuously deformed under ζ [7]. Thus, SBP is guaranteed to capture the global minimum and find the most accurate marginals in this setting. We will now empirically evaluate how this behavior extends to models with arbitrary random potentials.



(a) (b)

Figure 2: MSE and number of iterations for: SBP (blue), BP (magenta), and BP_D (green) with $\theta_i \sim \mathcal{U}(-0.5, 0.5)$ and (a) $J_{ij} \sim \mathcal{U}(0, \beta)$ (attractive model); (b) $J_{ij} \sim \mathcal{U}(-\beta, \beta)$ (general model).

Attractive models: We generate $L = 100$ models for every value of $\beta \in \{0, 0.5, \ldots, 5\}$ and sample the parameters according to $\theta_i \sim \mathcal{U}(-0.5, 0.5)$ and $J_{ij} \sim \mathcal{U}(0, \beta)$; i.e., overall we consider 1100 different models for each graph-structure. Note that BP is randomly initialized 100 times for every considered model. We compute the MSE for every value of β and visualize the mean and the standard deviation of the MSE as well as the number of iterations in Fig. 2a.

BP (magenta) converges rapidly for all graphs considered; hence, there is no additional benefit for BP_D (green) that only increases the number of iterations. SBP (blue) slightly increases the number of iterations as compared to BP but converges in fewer iterations than BP_D. We empirically observe that SBP consistently outperforms BP with respect to accuracy, specifically for models with strong couplings. These models exhibit multiple stable fixed points [7] such that, depending on the initialization, BP often converges to inaccurate fixed points.

General models: General models admit frustrated cycles and traditionally pose problems for BP and other methods that aim to minimize $\mathcal{F}_\mathcal{B}$. For every $\beta \in \{0, 0.5, \ldots, 5\}$ we consider $L = 100$ models and consider $\theta_i \sim \mathcal{U}(-0.5, 0.5)$ and $J_{ij} \sim \mathcal{U}(-\beta, \beta)$. We restrict the results to $\beta \leq 2$ on the grid graph because BP failed to converge for models with stronger couplings. Even for models where BP converged, SBP requires only slightly more iterations than BP and fewer than BP_D. Again SBP (blue) outperforms BP (magenta) and BP_D (green) on all graphs with respect to accuracy.

5 Discussion

We introduced an iterative algorithm that modifies belief propagation so that it converges towards accurate fixed points: self-guided belief propagation (SBP) is a simple and robust method that

gradually accounts for the pairwise potentials and guides itself towards a unique, stable and accurate solution.

Our empirical analysis demonstrates that SBP consistently improves the accuracy and provides more accurate marginals than BP for both attractive and general models. Moreover, SBP allows to approximate the marginals well on graphical models for which BP does not converge at all.

References

[1] E. Allgower and K. Georg. *Introduction to Numerical Continuation Methods*. Society for Industrial and Applied Mathematics, 2003.

[2] S. G. Brush. History of the Lenz-Ising model. *Reviews of Modern Physics*, 39(4):883, 1967.

[3] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.

[4] A. Globerson and T. S. Jaakkola. Convergent propagation algorithms via oriented trees. In *Proceedings of UAI*, 2007.

[5] T. Heskes. Stable fixed points of loopy belief propagation are minima of the Bethe free energy. In *Proceedings of NIPS*, volume 15, pages 359–366, 2003.

[6] G. E. Hinton and T. J. Sejnowski. Learning and relearning in boltzmann machines. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1(282-317):2, 1986.

[7] C. Knoll, D. Mehta, T. Chen, and F. Pernkopf. Fixed points of belief propagation – an analysis via polynomial homotopy continuation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[8] C. Knoll and F. Pernkopf. On loopy belief propagation – local stability analysis for non-vanishing fields. In *Proceedings of UAI*, 2017.

[9] C. Knoll and F. Pernkopf. Belief propagation: Accurate marginals or accurate partition function – where is the difference? In *Proceedings of UAI*, 2019.

[10] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

[11] T. Meltzer, A. Globerson, and Y. Weiss. Convergent message passing algorithms: a unifying view. In *Proceedings of UAI*, pages 393–401, 2009.

[12] O. Meshi, A. Jaimovich, A. Globerson, and N. Friedman. Convexifying the Bethe free energy. In *Proceedings of UAI*. AUAI Press, 2009.

[13] M. Mezard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

[14] J. M. Mooij and H. J. Kappen. On the properties of the Bethe approximation and loopy belief propagation on binary networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11012, 2005.

[15] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proceedings of UAI*, pages 467–475, 1999.

[16] J. Shin. Complexity of Bethe approximation. In *Proceedings of AISTATS*, pages 1037–1045, 2012.

[17] D. Sontag and T. S. Jaakkola. New outer bounds on the marginal polytope. In *Proceedings of NIPS*, pages 1393–1400, 2008.

[18] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.

[19] A. Weller and T. Jebara. Approximating the Bethe partition function. In *Proceedings of UAI*, 2014.

[20] A. Weller, K. Tang, T. Jebara, and D. Sontag. Understanding the Bethe approximation: When and how can it go wrong? In *Proceedings of UAI*, pages 868–877, 2014.

[21] M. Welling and Y. W. Teh. Approximate inference in Boltzmann machines. *Artificial Intelligence*, 143(1):19–50, 2003.

[22] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.